BASEMENT

MASSACHUSETTS
INSTITUTE of TECHNOLOGY
LIBRARIES

HD28
.M414
no1570-84

DEWEY

Micro-Production Functions Aren't Pretty:

Firm-Level and Industry-Level Specification

For Inputs and Outputs

by

Casey Ichniowski

WP # 1570-84 June, 1984

The author would like to thank Wayne Gray, Zvi Grilliches, Daniel McFadden and Robert Solow for their valuable comments on this paper.

FEB 1 3 1987

June 1984

Micro-Production Functions Aren't Pretty: Firm-Level and Industry-Level Specification For Inputs and Outputs

ABSTRACT

This study documents extreme variations in productivity within a panel of eleven firms in the same narrowly defined industry classification. Many of the sources of this variation were identified in field investigations of each plant. These investigations in turn allowed for the development of detailed specifications for inputs and outputs using data collected from the sites. The empirical estimates show that, irrespective of the precise functional form adopted, these detailed specifications, particularly those for output heterogeneity, are critical determinants of the performance of plant-level production functions. When the most detailed input and output specifications are used, 95% of the observed variation in plant production is explained. However, when the eleven firms are treated as an industry, less detailed specifications for inputs and outputs are shown to be more appropriate for explaining the variation in industry production.

Casey Ichniowski

I. Introduction

Economic theory treats the firm as something of a black box that efficiently transforms input into output. Since a firm's productivity is determined by available technology and conditions in the factor and product markets in the neoclassical framework, this "black box" view is not surprising. The view has lead some economists to believe that analysis and comparison of the operations of competing firms would reveal little of interest.[1] Add to this the problem of scarce data, and it is understandable why production functions are rarely estimated with establishment level data and are not intended to be direct representations of the operations of individual plants. Recently, however, micro-productivity studies have attracted greater attention from economists. These studies, often citing the large productivity residuals in aggregate studies that focus on a standard set of inputs, stress the need to consider other sorts of inputs[2] that can best be studied at the plant-level. For such studies, the traditional economic view provides very little guidance on how to develop a production function that can accurately account for the variation in input-output relationships at the plant-level. In this study, by analyzing a unique data set on eleven plants in the same four digit Standard Industrial Classification (SIC no. 2621-paper), I provide a guide on developing specifications for input and output variables in micro-level production functions.

The study is developed in five sections. Section II describes the plant data and the mix of econometric and field research used in this study. Section III describes the production processes in these paper mills. This section also

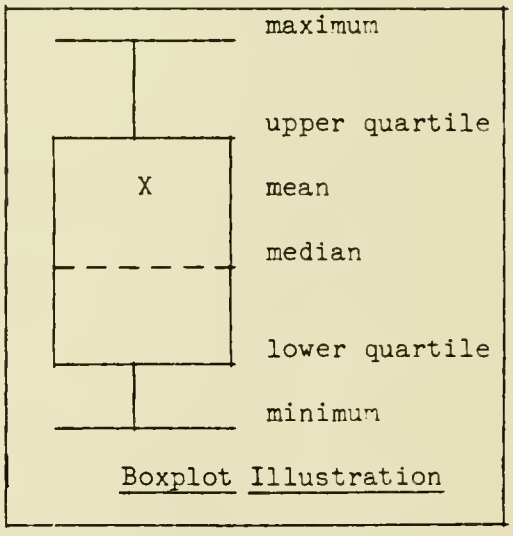
discusses how best to structure the available data to represent inputs and outputs for this panel of eleven plants. In Section IV, equations that incorporate the detailed input and output specifications are estimated. To gauge the contribution of these detailed specifications, the results from these equations are compared to results from equations using more conventional specifications for the variables. Section V illustrates whether the detailed specifications are necessary in more aggregate analysis by estimating the equations using data aggregated across the eleven mills.

By way of preview, the study yields four principal conclusions. Most importantly, the study flatly rejects the simple view that firms in the same narrowly defined industry classification are homogeneous configurations of equally productive inputs. Ouput, as well, is heterogeneous. The result is that productivity, by any metric, varies considerably in a narrowly defined industry sample. Second, a large number of sources of this variation were identified through field investigations of each mill. These investigations led to the development of detailed input variables and controls for output differences. Third, irrespective of the precise functional form adopted, the inclusion of these detailed specifications, particularly the controls for output heterogeneity, are critical determinants of the performance of the production function. For example, a simple Cobb-Douglas estimated with inputs defined as total labor hours, total value of capital, and total energy input is shown to be an extremely poor representation of plant production. However, expanding this model to include controls for output heterogeneity significantly improves the performance

of the Cobb-Douglas. The fourth conclusion indicates that there is still promise for the simple Cobb-Douglas without detailed output controls in industry analysis. The principal reason for the improved performance of the simple Cobb-Douglas at the industry level is that the controls for output heterogeneity are plant-specific and drop out of industry equations.

II. Sample and Methodology

The data in this study, monthly observations from January 1976 to September 1982, describe the operations of eleven paper mills. The initial hope was that with competing plants within a narrowly defined industry classification, many specification issues on inputs, output and functional form would be minimized. Figure 1 presents some initial evidence that lead to a reevaluation of these expectations. The figure shows the distribution of monthly values of the labor productivity index, tons per hourly manhour, for each mill in a "boxplot" (see illustration).



Boxplot Illustration

The average labor productivity for the whole sample is .247. Four mills consistently produce with higher rates of labor productivity; the other seven with below average labor productivity. There is also considerable within-mill labor productivity variation. Even for the most compressed distribution (mill 6, σ = .013), there is a 47.9% difference between the maximum and minimum values of the distribution. For the least compressed distribution, (mill 10, σ = .099), the maximum value of the labor productivity index is approximately three and one-third times greater than the minimum value.

Rather than assume that such extreme variation was due to differences in the intensities of complementary factors, I expanded the methodology to include

TONS
PER
MAN-HOUR
.60
.50
.40
.30
.20
.10
MILL 1
MILL 2
MILL 3
MILL 4
MILL 5
MILL 6
MILL 7
MILL 8
MILL 9
MILL 10
MILL 11
ALL MILLS

Figure 1

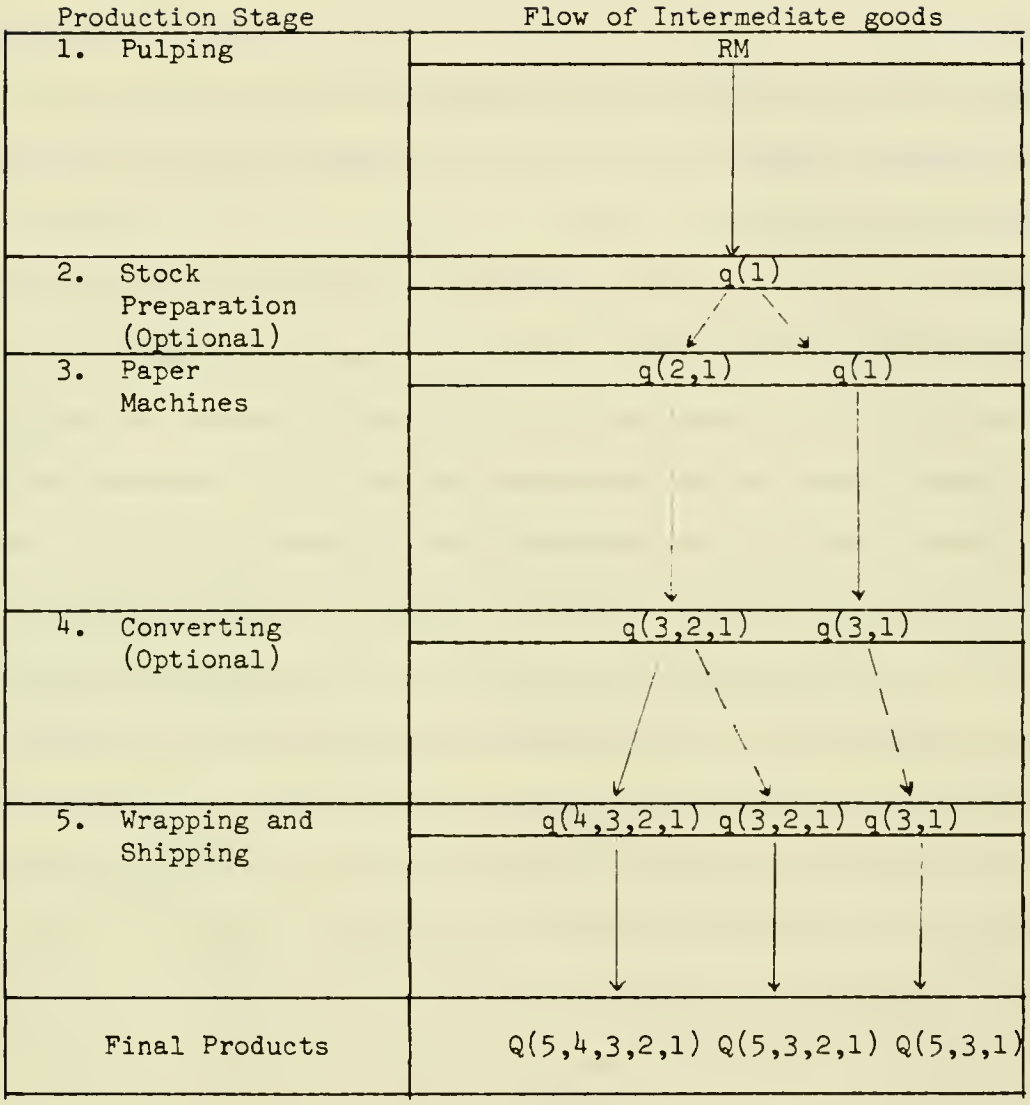
field investigations of each plant's production process with a mill engineer. Many theoretical issues were uncovered inside these supposedly similar firms: product differences across plants; multi-product technologies with issues of economies of scope; heterogeneous inputs particularly for capital; investments of varying vintage; limited within-plant subsititution possibilities once machinery is put in place. The next section considers how these issues can be addressed by using an understanding of the production processes to develop reasonable input and output specifications.

III. Plant Production Processes, Input-Output Data, and Model Specifications

To help understand why productivity indices for a set of firms in a narrowly defined industry would vary to the extent indicated in Section II, field interviews and tours of production processes were conducted in each of the eleven mills in the sample. The aim of the field research was to identify what was different from plant to plant and what might have changed during the seven year history of each plant to cause the variation in relative input-output ratios and configurations. A brief description of some of the indiosyncratic features of these mills provides a necessary background on how the available input-ouput data described below will be structured to account for productivity variations.

While paper-making is basically a continuous-flow process, the eleven mills can be represented as combinations of five stages: woodhandling and pulp production; stock preparation to convert dark brown pulp into the desired color and consistency; conversion of pulp into paper; additional converting operations such as sheeting and coating; and wrapping and shipping. Figure 2 represents the sequencing of these stages, and the flow of intermediate goods into final products. As illustrated in the diagram, one finding from the field research is that while the mills may be classified in the same four-digit industry, they are not in fact direct competitors producing identical output. Mills that do not have optional stages 2 and 4 can only produce final product Q(5,3,1) - large rolls of wrapped newsprint. When all five stages are present, Q(5,4,3,2,1) - coated or sheeted paper of higher quality than newsprint - is produced. While

Figure 2



Notes: $q_{(i)}$ refers to intermediate goods processed through stage i.

these mills can produce Q(5,3,1), the mill always operates its significant stage 2 and 4 capital investments to produce the higher-priced output. Similarly, Q(5,3,2,1) in Figure 2 represents large rolls of higher quality paper that has not been coated or sheeted.

In addition to these principal differences in output, the inputs within a given stage of production vary. For example, in stage 1, five different technologies were observed for producing pulp in the eleven mills: mechanical grinding of logs; sulfur chemical processing of wood chips; chemical processing in batch digesters; chemical processing in Kamyr digesters that convert wood chips continuously; and thermo-mechanical pulping which combines elements of mechanical and chemical processes. In addition, one mill buys market pulp rather than making its own. In the central paper-making process of stage 3, two principal paper machine technologies are used in the mills: a "twin-wire" and a "foundriner" process. The different technologies for a given stage across plants represent very different combinations of capital, labor, and energy inputs. While each stage of production across all mills appears to have fixed-factor characteristics, department-specific data could be used to estimate the types of substitution possibilities mapped out for a given stage of production across mills. Unfortunately, the input-output data available from company and plant sources, while extremely rich, are not detailed enough to develop department-or stage-specifc production functions. The discussion to follow considers how output and each major input can best be specified with available data to account for plant-level variation in production.

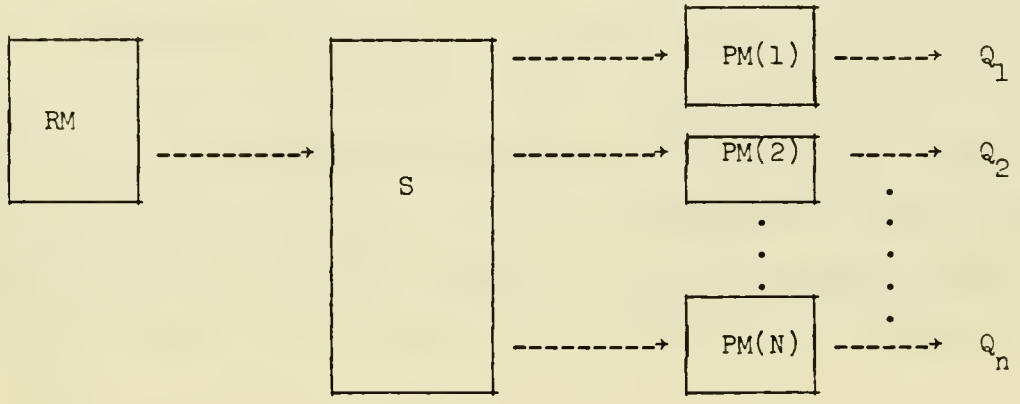
Output

The narrow focus of the sample did not insure homogeneous output. Two output indices are available for the study: tons produced and total sales. What adjustments to either of these indices are required to develop a reasonable index of the mills' output?

When product differences exist as in aggregate productivity studies, value added is the standard output index. To illustrate the data required to adjust net sales into a usable output index (i.e. to illustrate the ways that net sales differs from value added in this particular sample), simplify the production technology described in Section III to the multi-product process[3] shown in Figure 3. In Figure 3, raw materials (RM) are converted into a common stock(S) which can be processed through various combinations of additional stock treatment, paper machines, and converting operations (collectively represented as $PM_{(i)}$). These machines produce the array of final products, Q_1 to Q_n.

A simple model illustrates how the prices of the final products and inputs of the process can be used to weight the individual Q_i's to

Figure 3



produce a common output index. Each final product, q_i, requires a certain fraction, Θ_i, of the stock.

$$S = \sum^{n} \Theta_i \cdot Q_i \qquad \text{(Equation 1)}$$

The price of each final product, p_i, reflects in competitive markets the stock requirements and additional input requirements. For purposes of illustration, let the additional labor requirements (χ_i) represent all such input requirements. With an associated per unit labor cost of w, the product price can be written:

$$p_i = \Theta_i p_s + w\chi_i \qquad \text{(Equation 2)}$$

Summing across products, one obtains:

$$\Sigma p_i Q_i = p_s (\Sigma \Theta_i Q_i) + w\Sigma \chi_i Q_i \qquad \text{(Equation 3)}$$

Next, the stock requirements must be expressed. Let λ represent the labor requirements of one unit of stock and p_{RM} be the price of raw materials used to prepare the stock. The stock requirements equation becomes:

$$p_s (\Sigma \Theta_i Q_i) = p_{RM} \Sigma \Theta_i Q_i) + \lambda w (\Sigma \Theta_i Q_i) \qquad \text{(Equation 4)}$$

Combining 3 and 4, one solves for an expression that represents final products weighted by their prices:

$$\Sigma p_i Q_i = w[\Sigma (\chi_i Q_i) + \lambda \Sigma (\theta_i Q_i)] + p_{RM} (\Sigma \Theta_i Q_i) \qquad \text{(Equation 5)}$$

Under this simplified multi-product representation, one sees in equation 5 that the available net sales index is a function of a number of factors that vary with the specific product type Q_i. Furthermore, Figure 3 is not a particularly accurate representation of the production process. First, there exists no

common stock that is transformed by paper machines. The wood materials required for different papers vary greatly in quality and therefore price. To obtain an accurate value added index, net sales would have to be adjusted for these sorts of raw material differences. In addition, the markets in question are not necessarily competitive. Because of the weight of the commodity, the market for paper is regional. The isolated mills may enjoy some degree of market power as well as a degree of monopsonistic power in the labor market. Furthermore, unions for production workers may produce a situation of bilateral monopoly in the labor market. Wages therefore vary across mills and time periods. The price and wage adjustments to net sales would be extensive and far exceeds the capacity of even the rich data set collected for this study.

An alternative approach involves adjustments of the output variable, tons produced. In the discussion accompanying Figure 2, three main categories of paper produced in these mills were described: rolls of newsprint ($Q_{5(3,1)}$); rolls of white paper ($Q_{5(3,2,1)}$); and sheeted or coated white paper ($Q_{5(4,3,2,1)}$). One could adjust the tons variable by introducing dummy variables for the presence of optional stages 2 and 4. Both dummy variables should obtain a negative coefficient; that is, for a given level of inputs, a mill produces fewer tons of paper when it devotes inputs to either the stock preparation stage and/or the converting stage. Additionally, a dummy for the presence of stage 1 is required in the analysis for the mill in the sample that buys rather than makes its own pulp. This dummy too should obtain a negative coefficient. For the empirical work, then, the tons produced variable will be

used as the dependent variable along with three dummy variables which control for the presence of stages 1, 2, and 4.

Still, within any plant, there is a distribution of paper grades. The approach adopted for this study will be more appropriate if the distibution of paper grades produced in any one plant is relatively fixed. Any shift in the distibution of paper grades that does affect the input requirements should be accompanied by a change in the structure of the input variables developed below.

Capital Inputs

The mill tours revealed heterogenity in inputs as well as outputs. Capital stock, for example, is comprised of machinery in each stage of production, buildings, land, transportation equipment, office supplies, and pollution controls. Since the paper industry is one of the most capital intensive in the United States,[4] and since capital investments dictate the levels of other inputs, the specification of this input is critical. In this study, I construct a set of capital variables from the complete monthly inventories of each mill's assets. These inventories also give each asset's purchase price, and depreciation schemes based on engineering estimates of the life of each asset. For any month, there are as many as 15,000 individual assests in place in these mills. The task then is to transform this extensive list of assets into a set of useful capital variables.

Solow establishes the necessary and sufficient condition for collapsing two inputs into one in a production function: the marginal rate of substitution of one input for another must be independent of other inputs in use.[5] Applying

this principle here, I create nine categories of capital inputs. First, there are six variables that measure the capital in the five stages of the process described in Section III: wood handling and pulp production (stage 1); stock preparation (stage 2); paper machines (stage 3); converting operations (stage 4); and wrapping and shipping (stage 5). The argument for these aggregations is that virtually no substitution opportunities exist among the individual assets (e.g., pumps, screens, belt, wires, engines) in a given stage, regardless of the level of other inputs. For example, the individual engines, pumps, belts, or rollers that make up a coater are treated as indispensible components of one large machine. It is assumed that there are no substitution possibilities across these component assets of a coater regardless of the level of other inputs in the mill. Three other categories of capital inputs are developed: energy generation capital; pollution and recycling captial; and a miscellaneous category. Aggregation of the energy generation assets is motivated by the assumption that each asset in the category is an indispensible part of one large unit; however, energy or possibly labor may well be substituted for this capital. Therefore this category will not be combined with any other capital category. Pollution capital, purchased to meet various environmental standards, are unlikely to make the same contribution to output that other production machinery does; therefore this category will be kept separate from other categories. Finally, assets that I could not allocate to a particular category (approximately 13% of the value of all mill assets) are allocated to a separate category.

Specifically, these nine capital variables are created as follows: (1) each asset in each month is assigned to one of the nine categories; (2) the value of assets in a stage is calculated as the depreciated book value of those assets deflated by an industry-specific cost of captial. The depreciation scheme used in the inventories is straight-line depreciation allocated over the engineering life of each specific asset. The deflator[6] is intended to adjust for price inflation of otherwise equally productive machinery.

While the discussion of the production processes indicate some substition possibilities between capital in a given stage and some inputs (particularly purchased intermediate goods), a machine in one stage is only useful in that stage and not substitutable for machinery in other stages. For this reason, specifications will also be estimated that collapse the capital in the five stages into one measure (i.e., total value of direct production capital) as well as a total value of capital measure.

Scale

Economies of scale may exist in the mills. The capacity constraint of the mills is generally imposed by the capacity of the paper machines. A set of dummies describing the number of paper machines in the mill is added to control for possibility of scale economies. In addition, since a paper machine's capacity varies with vintage, depreciation and deflation of the paper machine's value may also help to adjust for these sorts of vintage effects. While these variables will not be a perfect control for scale, they should provide some measure of control over the possibly critical issue of paper machine capacity.

Labor

Labor input is measured by total production manhours. An accurate salaried manhours variable was unavailable. Additionally, the production manhours variable has two principal components in these mills: operating and maintenance labor. Separate variables for these components of the labor input were not available. Finally, data on the labor input associated with each stage of the production process were also unavailable.

Energy

The energy input is measured as the total number of BTU's used per month in a mill. One drawback of this variable is that it is not an "efficient" BTU measure. More BTU's may be used in a plant simply because certain sources of energy provide BTU's more efficiently. As with the labor input, the energy input is not broken down by usage in each stage of production.

Raw Materials

Detailed raw material data were unavailable. However, raw material requirements are by and large dictated by the final product desired. Therefore, if the capital dummies controlling for product and process variations are adequate, this omission should not greatly affect the equations ability to account for productivity variation.

Econometric Specification

The input and output specifications above are described as desirable elaborations that should be incorporated in the production analysis regardless of specific functional form chosen. In the next section, specifications will be

estimated with the detailed input and output variables and then compared to results obtained from specifications with lesser degrees of detail. Perhaps the simplest specification that incorporates all of the details for the specifications of inputs and outputs would be an equation of the following form:

$$Q = B_0 + B_1 \cdot \vec{KD}_{1-3} + B_2 \cdot \vec{K}_{1-9} + B_3 \cdot \vec{SC} + B_4 \cdot L + B_5 \cdot E \qquad \text{(equation 6)}$$

where Q = the natural logarithm of output

$\vec{KD}_{1-3}$ = a vector of three dummies for the presence of stages 1, 2, and 4 that control for heterogeneity of products and heterogeneity in processes for providing similar products

$\vec{K}_{1-9}$ = a vector of nine capital value variables (measured in absolute dollars due to the absence of capital in certain stages)

$\vec{SC}$ = set of dummies for the number of paper machines as a measure of scale

L = log of total production manhours

E = log of total BTU's used.

While equation 6 may provide a reasonable local approximation that accounts for variation in production for the observed range of input configurations, certain assumptions embodied in the functional form clearly make it unreasonable to extend the interpretation of coefficients outside the range of observed values. For example, individual inputs do not seem to change in isolation; higher levels of capital in one stage might have to be associated with higher levels of capital in the other stages before increases in output are

realized. To see if this relatively simple equation 6 specification using the detailed input variables and output controls provides a reasonable local approximation of how production varies over the observed range of input configurations, this model is estimated in the next section. After describing these estimates, they will be compared to results from a Cobb-Douglas model with more conventional input specifications. The Cobb-Douglas model is then expanded to incorporate additional degrees of detail contained in the equation 6 model to see if the added details serve to improve this functional form for the production function. The hope is that the detailed input and output specifications improve the performance of the production model regardless of its specific functional from.

V. Estimation of Plant-Level Equations

Table 1 represents the coefficients obtained from estimating the model in Equation 6. Despite the lack of stage-specific data on all inputs and the potential problems associated with the assumptions embodied in this simple functional form, all coefficients are quite consistent with the intuition behind the description in the previous section. All of the capital dummy variables for stages of the production process (lines 1a, 2a, and 4a) are negative. Production in tons is significantly lower relative to other mills when the paper being produced is: converted paper (line 4a); white paper or other paper requiring significant forms of stock preparation (line 2a); or paper produced from roundwood or chips rather than from market pulp (line 1a). A regression equation was also estimated that expressed the natural logarithm of hourly manhours as a function of the capital variables to provide a better understanding of the reasons for such lower tonnages of paper. This supplementary analysis revealed that producing converted paper requires approximately 20.2% additional hourly manhours, while those plants with significant stock preparation, mostly the white paper mills, require 7.6% more hourly manhours. It is for these reasons that the dummy capital variables are critical control variables in the productivity model.

The coefficients on the variables for production capital (woodhandling - line 1(b1); pulping - line 1(b2); stock preparation - line 2(b); converting capital - line 4 (b); and wrapping and shipping - line 5(a)) are all positive and significant at conventional levels. More capital in any of these stages is

Table 1: Complete Set of Regression Coefficients from General Productivity Equation[a]
(N=696)

INDEPENDENT VARIABLES	Dependent Variable: ln TONS of Paper
1. Wood Handling/Pulping Capital	
a. dummy for stage	-.140***
	(.048)
b.(1) value in woodhandling	.089 E-4**
	(.051 E-4)
b.(2) value in pulping	.115 E-4***
	(.035 E-4)
2. Stock Prep Capital	
a. dummy for stage	-.096***
	(.032)
b. value in stage	.072 E-4***
	(.026 E-4)
3. Paper Machine Capital	
a. 2PM dummy	-0-
b. value of 2PM's	.036 E-4*
	(.026 E-4)
c. 3PM dummy	-.478***
	(.034)
d. value of 3PM's	.016 E-4
	(.021 E-4)
e. 4PM dummy	-.793***
	(.039)
f. value of 4PM's	.071 E-4***
	(.012 E-4)
g. 5 or more PM dummy	-1.359***
	(.208)
h. value of 5 plus PM's	.458 E-4***
	(.186 E-4)
4. Converting Capital	
a. dummy for stage	-.291***
	(.038)
b. value in stage	.766 E-4***
	(.124 E-4)

Table 1 (Con't)

INDEPENDENT VARIABLES	Dependent Variable: ln TONS of paper TOTAL MANHOURS
5. Wrapping/Shipping Capital	
a. value only	.325 E-4***
	(.119 E-4)
6. Energy/Recycling Capital	
a. value only	-.067 E-4***
	(.016 E-4)
7. Pollution Capital	
a. value only	-.012 E-4
	(.022 E-4)
8. Land/Office/Other Capital	
a. value only	-.252 E-4***
	(.017 E-4)
9. Energy - ln BTU's Used	.356***
	(.021)
10. Labor - ln Hourly Manhours	.612***
	(.026)
R^2	.955
SSR	10.869

[a] - standard errors in parentheses

* - significant at the .10-level, one-tailed test

** - significant at the .05-level, one-tailed test

*** - significant at the .01-level, one-tailed test

consistently associated with higher production. Again, these capital inputs probably do not move in isolation. Higher levels of capital in one stage may only exist in combination with appropriate increases in capital in other stages. Still, higher levels of capital (regardless of how the increase is split among the stages in these observations) are associated with higher levels of output. Similarly the coefficients on the paper machine value variables all show positive coefficients (lines 3b, d, f, and h). Using the number of paper machines as an indicator of scale, one observes that the coefficients on the paper machine dummy variables become more negative as the number of paper machines increases.

The coefficients in lines 6 through 8 indicate the importance of separating the "energy generation," "pollution control," and "land buildings, and other" capital from production process capital. None of this capital stimulates productivity as does other capital. The coefficients on the capital variables in the "energy generation" and "land" categories are significantly negative. Clearly, the assumption of a completely homogeneous capital stock would have been inappropriate in this setting.

Finally, the coefficients on the labor and energy variables are also both significant and positive. Overall, the model accounts for nearly 96% of the total variance in production. To demonstrate the performance of the model further, the residuals from equation 6 are grouped by mill and displayed in boxplots in Figure 4. The large advantages observed for mills 1, 7, 10, and 11 in Figure 1 are no longer evident. The distribution for each plant is generally

.20
.00
-.10
-.20
MILL 1
MILL 2
MILL 3
MILL 4
MILL 5
MILL 6
MILL 7
MILL 8
MILL 9
MILL 10
MILL 11

Figure 4

Table 2: Cobb-Douglas Specifications with Plant-Level Data[a]
[Dependent Variables: ln Tons of Paper; N=696]

	(1)	(2)
1. ln (total value of capital)	-.010 (.017)	.040*** (.010)
2. ln (hourly manhours)	.128*** (.029)	.771*** (.026)
3. ln (BTU's used)	.595*** (.027)	.465*** (.020)
4. Dummy for stage 1	____	-.791*** (.044)
5. Dummy for stage 2	____	-.184*** (.029)
6. Dummy for stage 4	____	-.262*** (.025)
7. Three dummies for number of paper machines	No	Yes
R^2	.648	.904
MSE	.122	.033

[a] - standard errors in parentheses

* - significant at the .10-level, one-tailed test

** - significant at the .05-level, one-tailed test

*** - significant at the .01-level, one-tailed test

Cobb-Douglas model. The two principal differences between the Cobb-Douglas (Table 2, column 1) specification and the Equation 6 specification (Table 1) are the dummies for production stages and number of paper machines, and the expansion of the capital variable into detailed components.

To see how much of the difference in performance is due to the dummy variables, the Cobb-Douglas specification is expanded to include these variables. These results are shown in column 2 of Table 2. The performance of the Cobb-Douglas improves in a number of respects. First, the capital coefficient is now significant and positive. The coefficients on labor (line 2) and on energy (line 3) are more similar to the parameter estimates obtained from the equation 6 model (Table 2, lines 9 and 10). Also the dummies for stages of production (lines 4, 5, and 6) are significantly negative although different in magnitude than those in Table 1. The R^2 after introducing the dummy increases to .904 and the mean squared error reduces to .033. These dummies which help control for heterogeneity in output and in production processes (make vs. buy pulp in stage 1) are clearly important controls regardless of precise functional form chosen.

Some additional increases in explanatory power and reductions in mean squared error are achieved from expanding the capital input into more detailed components. A comparison of the summary statistics of the equation in Table 1 and the equation in column 2 of Table 2 illustrates this point. An alternative illustration of the contribution made by the detailed capital variables is to estimate mill-specific equations.[7] To adapt the general model for plant-

centered on zero, the average residual value. However, mill 7 and especially mill 11 appear to be relatively productive mills given the measured levels of inputs, while mill 10 appears to be relatively unproductive.

While the model eliminates the large cross-plant productivity differentials and provides a high level of explanatory power, the performance of the model can be compared to alternative specifications without the same level of detail in input and output specifications. As a popular empircal alternative, consider a conventional Cobb-Douglas production function:

$$Q = AK^{\alpha 1}L^{\alpha 2}E^{\alpha 3} \qquad \text{(Equation 7)}$$

where K, L, and E are total values of capital, labor and energy respectively and A, α_1, α_2, and α_3 are the parameters free to vary. After a log transformation, equation 7 becomes:

$$\ln Q = \ln A + \alpha_1 \ln K + \alpha_2 \ln L + \alpha_3 \ln E \qquad \text{(Equation 8)}$$

Estimating equation 7 by OLS produces the parameter estimates in column 1 of Table 2. Interpreting these coefficients in the standard way would have been very misleading. The estimated coefficients on labor and energy are very different from those obtained from the equation 6 model (Table 1, line 9 and line 10). The capital coefficient is particularly misleading. The coefficient is not significantly different from zero and the point estimate is actually negative. This is clearly a nonsensical result in this highly capital intensive industry. The R^2 for the Cobb-Douglas model is .648 as compared to .955 for the equation 6 model. Using the mean-squared error as a criterion to judge the performance of these models (S^2 = SSR/degrees of freedom), the S^2 for the equation 6 model is .0162 as opposed to .122 for the Cobb-Douglas model. The two prin-

specific equations, all stage-of-production dummies are dropped from the equation since no plant changed its basic production process during this period. In this way, mill-specific equations will not contain dummies for these cross-plant product and process differences. Paper machine dummies only appear for three mills that changed the number of paper machines in operation during the seven-year period. Detailed capital variables for any particular stage will appear for the plants that have the stage.

After making the necessary changes, the detailed model is estimated separately for each mill. The average R^2 for the mill-specific equations is .781. The minimum R^2 is .550; the maximum is .959. Again, the performance of these equations can be evaluated against a Cobb-Douglas alternative without the added detail in the specification of capital inputs. When the Cobb-Douglas production functions are estimated separately for each mill the average R^2 is .598. The average S^2 for the Cobb-Douglas equation is .0092, while for the detailed production equations it is only .0053. The complete set of R^2's and S^2's is given in the Appendix to this chapter in Table A. In all cases, the Cobb-Douglas model produces a lower R^2 and a higher S^2.

While these plant-specific equations may be more appropriately estimated by linear programming because of the fixed-factor characteristics observed in field investigations, the empirical results in this section taken as a whole seem to warrant two principal conclusions. First, the relatively poor performance of the simple Cobb-Douglas specification at the micro-level seems in large part due to important cross-mill differences in products and processes. The control variables for these differences are necessitated by the nature of the

dependent variable. In the absence of a value added dependent variable that might inherently control for differences in the degree to which products are processed, the dummy variables are critical controls in the model. Second, conventional input specifications, particularly the inclusion of pollution controls and other non-production capital in the capital stock, also are not entirely satisfactory.

These two difficulties are not so much problems with the functional form of the Cobb-Douglas (or conversely advantages of the functional form of the equation 6 model). Rather, the differences are in the specification of the input and output variables within the functional form. That is, even with more flexible functional forms such as the CES and translog, one would still want to incorporate the product and process dummy variables and the added information provided by the detailed capital variables in the more flexible functions as well.[8]

V. Estimation of Industry-Level Equations

These issues related to input and output specifications, while critical in accounting for plant-level variations in production, may not be serious issues when estimating industry-level equations. Houthakker, in summarizing his theorem on how a Cobb-Douglas form results from the aggregation of fixed-factor plant equations, suggests that his theory should be extended "... to take more account of the fact that virtually all industries produce many different products, a compliction which cannot be satisfactorily dealt with by [his] present approach."[9] He argues that to test the empirical relevance of this extension to his theory, "data would have to be gathered on the distribution of input-output ratios between firms... [so as not] to neglect the variability in production possibilities between firms."[10] The basic principle of Houthakker's theorem has already been incorporated in the empirical work in this paper.[11] However, nearly thirty years after his suggestions for extending his theorem, this is believed to be the first attempt to illustrate the empirical relevance of his propositions concerning the aggregation of disparate intra-industry input-output configurations.

The analysis now considers the performance of equations after aggregating the data across the eleven mills. Table 3 presents these empirical results. In all specifications, the dummies for plant-specific product and process differences drop out since they sum to constants in industry-level data. Column 1 presents the coefficients from the model with the most detailed specification for the capital input. In this specification, the paper machine dummies take on

Table 3: Coefficients from Industry-Level Production Functions
Dependent Variable: ln Tons of Paper
(N=79)

Independent Variables	(1)	(2)	(3)
1. Labor			
a. ln production manhours	.793*** (.078)	.494*** (.088)	.527*** (.086)
2. Energy			
a. ln BTUs	.231*** (.069)	.243*** (.094)	.220*** (.090)
3. Capital Variables (in ln units)			
a. total value	———	.185*** (.037)	———
b. total value of production machinery	———	———	.173*** (.031)
c. woodhandling	.130*** (.051)	———	———
d. pulping	.375*** (.062)	———	———
e. stock preparation	.261*** (.054)	———	———
f. paper machines	- .081* (.062)	———	———
g. converting operations	.021* (.015)	———	———
h. wrapping/shipping	- .061 (.077)	———	———
i. energy generation	- .242** (.119)	———	———
j. pollution controls	- .153*** (.054)	———	———
k. other capital	- .305*** (.092)	———	———
l. paper machine dummies	yes	no	no
R2	.992	.967	.970
SSR	.141	.546	.509
MSE	.002	.007	.006

a - standard errors in parentheses
*** - significant at the .01 - level, one-tailed test
** - significant at the .05 - level, one-tailed test
* - significant at the .10 - level, one-tailed test

different values than they do in plant equations, now ranging from 29 to 41 machines at the industry-level over this period. The detailed model accounts for nearly all of the variation in industry output. However, the coefficients on paper machine capital (line 3f) and wrapping and shipping capital (line 3h) no longer have the expected positive sign. Coefficients on the capital outside of the prodution process (lines 3i,j,k) are still significantly negative.

Column 2 presents the estimates from the Cobb-Douglas production function using only the simple input specifications for total values of capital, labor, and energy. The capital coefficient is no longer negative as it was in the plant equation (Table 2, column 1), but rather significant and positive. The energy coefficient is similar to the coefficient in column 1. The manhours coefficient is significantly positive, yet smaller in magnitude than its column 1 counterpart. Furthermore, the inclusion of nonproductive capital (lines 3i, j, k) in the capital variable does not greatly affect the estimated coefficients. In column 3, the Cobb-Douglas specification is re-estimated using only the prodution process capital for the capital measure. The three input coefficients are still significantly positive and similar to the magnitudes of the column 2 coefficients.

These differences in coefficients when going from plant-level to industry-level data are one indication that simpler input and output specification are more appropriate at the industry-level. Another striking indication of the improved performance models with simpler input specifications at the industry-level is the amount of variation explained by the models with the less detailed

input specifications. The R^2 for the column 2 and column 3 Cobb-Douglas specification are approximately 97%--a marked increase over the explanatory power of the plant-level Cobb-Douglas equations. The improved performance of the Cobb-Douglas with simpler input specifications is perhaps not surprising. In plant-level equations, the product and process dummy variables for stages of production improves the performance of the simple Cobb-Douglas specification dramatically (Table 2, column 2 vs. Table 2, column 1); these variables drop out of industry-level equations.

The disappearance of the dummy variables for product and process differences at this level of analysis is surely part of the reason for the improved performance of the Cobb-Douglas equation. An additional reason why R^2's may increase when going from plant-level to industry-level data has been suggested by Grunfeld and Grilliches. They illustrate how larger multiple correlation coefficients can result "... when we aggregate...because within each group (industry) the individuals (firms) are likely to be more alike with respect to independent variables than with respect to disturbances."[12] For this panel of firms, however, it does not appear to be the case that input variables at the plant-level (using the simple input specifications) have a tighter distribution than do the corresponding error terms.[13] This factor, then, would not be responsible for the higher R^2's for the simple Cobb-Douglas model obtained after aggregation. In short, the Cobb-Douglas equation with the less detailed specifications for inputs appears to be a much more appropriate specification for industry-level data than for plant-level data. The principal reason for this

seems to be that product and process dummy variables need to be included in plant-level equations; such plant-specific variables, however, drop out of industry-level equations.

VI. Conclusion

The empirical results in this study illustrate several problems that are likely to arise in conducting micro-level productivity research. The paper also illustrates how field research or interviews can be used to provide solutions to these problems. Specifically, while research at the micro-level might be the most appropriate way for identifying the types of factors that lead to more or less efficient or productive organizations, modelling the inputs and outputs in the production process may be particularly difficult at this level. Most importantly, a skeptical eye should be used in considering micro productivity studies that assume that analyzing a sample of firms in a narrowly defined industry insures homogeneity of outputs or production technologies. Value added indices, the conventional method to control for output heterogeneity, can be extremely difficult to construct accurately as is the case in this sample. Physical output, while more easily measured, is shown to require adjustments before it can be used as a valid cross-plant output index.

The unique data in this study are also used to consider questions on aggregation that have remained unanswered in the productivity literature for some time. In response to some questions posed in the conclusions of important articles on producton over twenty years ago: (1) As Houthakker suggested, smoother functional forms seem to be reasonably appropriate even if plant production is characterized by fixed-factor technology, as long as cross-plant variation in these technologies is observed;[14] (2) as Grunfeld and Grilliches suggest, aggregation is not necessarily bad and has the advantage of eliminating

plant-specific factors from consideration;[15] and (3) as Arrow, Chenery, and Minhas suggested, inputs can also be usefully subdivided, such as the capital in different stages of production or the distinction between maintenance, operating and salaried labor.[16]

Perhaps, the most important conclusion to be drawn from this study is the reminder to the empirical researcher to develop as reasonable a framework as possible for productivity research given the level of analysis and the nature of the data. In this study the detailed model of plant production, which accounts for over 95% of the total variation in production, is intended to provide some guidance and illustrations of some of the specification issues for inputs and outputs that can arise even within a narrowly defined industry group.

Finally, the framework developed in this study is being expanded to consider less conventional "inputs", such as industrial safety, grievance activity, and strike activity, as possible determinants of a firm's performance. While a large portion of the variation in productivity in this sample is accounted for by the models in this study, a large amount of within-plant variation remains. The role of these less conventional inputs may also add to an understanding of what determines productivity inside the microeconomic "black box."

APPENDIX

Table A: R^2's and S^2's from Detailed Production Models and Cobb-Douglas Production Models

Sample	Detailed Production Model		Cobb-Douglas Production Model	
	R^2	S^2	R^2	S^2
1. pooled sample of 11 mills	.955	.0161	.648	.1216
2. mill 1	.774	.0034	.701	.0034
3. mill 2	.907	.0021	.625	.0077
4. mill 3	.723	.0066	.462	.0109
5. mill 4	.550	.0050	.163	.0084
6. mill 5	.768	.0034	.622	.0051
7. mill 6	.655	.0030	.517	.0035
8. mill 7	.730	.0020	.522	.0030
9. mill 8	.959	.0087	.929	.0133
10. mill 9	.939	.0055	.916	.0068
11. mill 10	.787	.0127	.438	.0295
12. mill 11	.800	.0061	.681	.0090

FOOTNOTES

1. For a discussion of the constraints imposed by the assumptions on the neoclassical framework, see Nelson, Richard R., "Research on Productivity Growth and Productivity Differences: Dead Ends and New Departures," Journal of Economic Literature, Vol. 19, (September 1981), p. 1037.

2. For some examples of these sorts of studies, see Nelson, pp. 1033-1045.

3. For a more general theoretical development of specification issues in multi-product technologies, see Hall, Robert," The Specification of Technology with Several Kinds of Output," Journal of Political Economy, (May 1973), pp. 878-892.

4. Data for 450 four digit standard industrial classifications in the manufacturing sector are available from U.S. Bureau of the Census, Annual Survey of Manufactures. Capital expenditures per production worker in the paper industry exceed the capital-labor ratio for all but 29 of these 450 manufacturing industries. Most of the other industries that are more capital intensive are in metal or petro-chemical industries. These statistics are for 1976, the first year used in the productivity analysis in Section V of this study.

5. Solow, Robert M., "The Production Function and the Theory of Capital," Review of Economic Studies, Vol. 23, No. 55, (1955-1956), pp. 101-108.

6. Capital price indices are taken from Economic Report of the President (Washington, D.C.: U.S. Government Printing Office, (1981), p. 229. Table B-57, Producer Price Indexes for Major Commodity Groups - Machinery and Equipment.

7. These equations can also be used to test the importance of standard time series concerns related to serial correlations in the panel data. Durbin-Watson statistics revealed no evidence of AR-1 correlation in the errors; more detailed inspection of residual from these plant specific equation (or from the equations 6 model when plotted for each mill over time) revealed no clear pattern in residuals. The inclusion of month dummies in the equation 6 model does indicate the presence of some seasonality of production, but these dummies do not affect other coefficients to any appreciable extent.

8. The important and related issue of substition possibilites that are mapped across plants that is not addressed in this paper is the subject of current research in progress. This research will consider a number of approaches for entering the detailed input and output specification within the Cobb-Douglas, CES, and trans-log functions. Substition parameters obtained using different functional forms with similar input specifica-

tions are incorporated within different functional forms. For the development of the CES form, see Arrow, K.J., H.B. Chenery, B.S. Minhas, and R.M. Solow, "Capital Labor Substitution and Economic Efficieny," Review of Economics and Statistics, Vol. 63, No. 3, (August 1981), pp. 225-250; for the translog form, see Christensen, L.R., D.W. Jorgensen, and L.J. Lau, "Conjugate Duality and the Transcendental Logarithmic Production Function," (abstract), Econometrica, Vol. 39, No. 4, pp. 255-256.

9. Houthakker, H.S., "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," Review of Economic Studies, Vol. 23, (1955-1956), p. 31.

10. Ibid., p. 31.

11. Whether with industry-level data, or with panel data across plants, the intuition on why a smooth curve can be adopted even if plant production has fixed factor characteristics is similar in both cases: if plant input configurations are clustered at the vertices of different discontinuous curves, a smoother curve will be mapped out when considering data across the plant-level observations.

12. Grunfeld, Y. and Zvi Grilliches, "Is Aggregation Necessarily Bad?" Review of Economics and Statistics, Vol. 62, No. 1, (February 1960), p. 4.

13. A simple test estimates each of the inputs and the errors from the simple Cobb-Douglas as a function of a complete set of month dummies. The variance

explained (mean square error) is, in fact, higher (lower) when the residuals are used as the dependent variable in this model than when the labor, capital, and energy inputs serve as the dependent variables. A more complete accounting of errors in aggregation is developed by Boot, J.C.G., and G.M. De Wit, "Investment Demand: An Emprical Contribution to the Aggregation Problem," International Economic Review, vol. 1, no.1, (January, 1960).

14. Houthhakker, pp. 27-31.

15. Grunfeld and Grilliches, pp. 1-13.

16. Arrow, Chenery, Minhas, and Solow, p. 247.

5804 644

Date Due

Lib-26-67

BASEMENT